Least squares

argmax/min invariant to 1/2 but it is useful in the computations: $\frac{\partial E}{\partial \tilde{W}} = 0$ where E is quadratic

$$\underbrace{\begin{pmatrix} \vdots \\ \text{---} \\ \vdots \end{pmatrix}}_{\tilde{X}\tilde{W} \text{ prediction}} - \underbrace{\begin{pmatrix} \vdots \\ \text{---} \\ \vdots \end{pmatrix}}_{T \text{ ground truth}} = \text{error}$$

Column view

$$A^T \cdot A = \begin{bmatrix} c_1^T \\ \vdots \\ c_N^T \end{bmatrix} [c_1 \cdots c_N]$$

in the diagonal you find the squared norm of each column

Minimize the sum-of-squares error function $E(\tilde{W}) = \frac{1}{2} Tr \{ (\tilde{X}\tilde{W} - T)^T (\tilde{X}\tilde{W} - T) \}$

closed-form solution $\tilde{W}^* = \boxed{(\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T} T$ pinv(x) · T in matlab/numpy

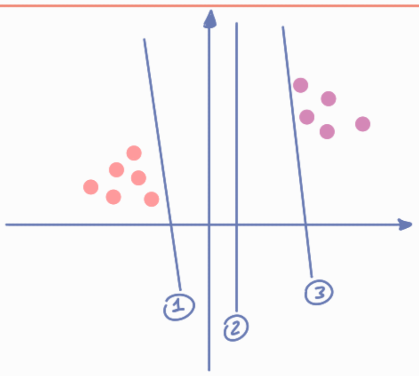
pseudo inverse of X

Once we have $\tilde{W}^*$, to classify a new instance:

$$y(x) = \tilde{W}^T \tilde{x} = \begin{pmatrix} y_1(x) \\ \vdots \\ y_N(x) \end{pmatrix} \longrightarrow K = \underset{i \in [1, \cdots, K]}{\operatorname{argmax}} \begin{pmatrix} y_1(x) \\ \vdots \\ y_N(x) \end{pmatrix}$$

Problem: not robust to outliers a.k.a. points in the dataset that possibly comes from a different distribution

Outlier: an outlier is a data point that differs significantly from other observations



① possible solution with small ρ

② not a possible solution with small ρ

③ possible solution with small ρ

Solution ② should be instead preferable since it better classifies instances not in the dataset.